DOCUMENT RESUME

ED 137 394 TM 006 200

AUTHOR Crehan, Kevin D.; And Others
TITLE Developmental Aspects of Test-Wiseness.
PUB DATE [Apr 77]
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Elementary Secondary Education; High School Students; Instructional Program Divisions; *Longitudinal Studies; Reliability; Response Style (Tests); Secondary Education; Sex Differences; *Statistical Analysis; Test Reliability; *Test Wiseness

ABSTRACT

Longitudinal studies of test wiseness (TW) were conducted to determine: (1) the relationship between TW and grade level, (2) the relationship between TW and sex, and (3) the stability of TW. Aspects of TW observed included stem cue and specific determiner identification and usage and the elimination of similar and absurd options. Subjects were observed three times at two year intervals in grades 5, 7, and 9; 6, 8, 10; 7, 9, 11; and 8, 10, 12. Results showed low to moderate internal consistency for the 16 item TW scale and test-retest correlations of approximately the same magnitude as KR20 indicating TW was somewhat stable over the intervals observed. Sex by year MANOVA trend analysis showed no sex effects and no sex by year interactions. A significant linear trend was observed over all four-year intervals with a quadratic trend appearing in the 7, 9, 11 sample. (Author/RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED137394

DEVELOPMENTAL ASPECTS OF TEST-WISENESS

Kevin D. Crehan
University of Nevada, Las Vegas

Leon J. Gross
University of Illinois

Roger A. Koehler
University of Nebraska

Malcolm J. Slakter
State University of New York at Buffalo

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Abstract

Longitudinal studies of test-wiseness (TW) were conducted to determine: (a) the relationship between TW and grade level, (b) the relationship between TW and sex, (c) the stability of TW. Aspects of TW observed included stem cue and specific determiner identification and usage and the elimination of similar and absurd options. Subjects were observed three times at two year intervals in grades 5, 7, and 9 (n=75); 6, 8, 10 (n=76); 7, 9, 11 (n=73); and 8, 10, 12 (n=64).

Results showed low to moderate internal consistency for the 16 item TW scale and test-retest correlations of approximately the same magnitude as KR20 indicating TW was somewhat stable over the intervals observed. Sex by year MANOVA trend analysis showed no sex effects and no sex by year interactions. Significant ($P<.05$) linear trend was observed over all four-year intervals with a quadratic trend appearing in the 7, 9, 11 sample.

TM006 200

DEVELOPMENTAL ASPECTS OF TEST-WISENESS

Kevin D. Crehan
University of Nevada, Las Vegas

Leon J. Gross
University of Illinois

Roger A. Koehler
University of Nebraska

Malcolm J. Slakter
State University of New York at Buffalo

Millman, Bishop, and Ebel (1965) define test-wiseness (TW) as "a subject's capacity to utilize the characteristics and formats of the test and/or test taking situation to receive a high score." Ebel (1972) suggests that TW is a concern when the examinee is faced with novel item types or when common item faults are present.

Previous research has shown that TW is not highly related to cognitive ability (Diamond and Evans, 1972) and that TW can be learned (Gibb, 1964; Moore, Schutz, and Baker, 1966; Slakter, Koehler, and Hampton, 1970b).

The relationships among TW, grade level and sex have been investigated in cross sectional (Slakter, Koehler, and Hampton, 1970a) and longitudinal studies (Crehan, Koehler, and Slakter, 1974). The longitudinal studies (Crehan et al., 1974) involved observation of four aspects of TW for students in grades 5 through 9 at time one and repeated observations of the same students two years later in grades 7 through 11. Results indicated: (a) no sex by grade interaction, (b) increases in TW over all intervals

Paper presented at the annual meeting of the American Educational Research Association, New York, N.Y., April, 1977.

except the grade 9 to 11 interval, (c) little evidence for sex differences in TW, and (d) TW remained relatively stable over the intervals observed. The present research reports the results of a third observation of TW following an additional two year interval.

Method

Ss were all available public school students in grades 9 through 12 who had been observed on both the first and second occasions. The maximum possible sample size was 391 and matches were obtained for 288 Ss.

The measurement instruments included the same 16 TW items embedded in 28 legitimate items which varied to suit the particular grade level. The TW items are designed to measure four aspects of TW from among those listed by Millman et al. (1965). Four items each are designed to measure the following TW behaviors:

1. recognize and select an option which resembles an aspect of the stem (stem-option),
2. recognize and eliminate those options which contain specific determiners, e.g., always, never, only (specific-determiners),
3. eliminate those options which are known to be incorrect and select from among the remaining options (absurd-options), and
4. eliminate similar options and select from among the remaining alternatives (similar-options).

The first two TW behaviors involve cue using strategies and the second two TW aspects involve deductive reasoning. With the exception of the aspect of TW being observed, the items were content free. That is, the keyed response could <u>not</u> be determined from knowledge of the nominal subject matter. Below are presented examples of items designed to measure each aspect of TW and the basis for identification of the keyed response.

1. Stem-Option:

The flying-spider is known for its ability to:

1. blend in with its surroundings.
* 2. glide through the air.
3. kill its prey with poison.
4. make very large webs.

Since the flying-spider does not exist, the item cannot be answered from knowledge of content. However, the test-wise examinee is expected to recognize the correspondence between the stem-cue "flying" and the option "glide through the air."

2. Specific-Determiners:

The Japanese game of paduki:

1. can only be played by the Imperial Family.
2. can never be played by teams.
* 3. is sometimes played indoors.
4. is played at every celebration.

Since the game of paduki does not exist and very few absolutes are universally true, the test-wise examinee is expected to eliminate options containing specific-determiners and select from among the remaining options.

3. Absurd-Options:

The greatest advantage of using slent in the manufacture of steel is that slent makes the steel:

* 1. stainless
2. transparent
3. heavy
4. bulky

The test-wise person is expected to eliminate options 2, 3, and 4 from consideration and identify option 1 as the only plausible alternative.

4. Similar-Options:

When Bestor crystals are added to water:

1. heat is given off.
2. the temperature of the solution rises.
3. the container becomes warmer.
* 4. the solution turns blue.

Options 1, 2, and 3 can be eliminated since they imply a similar result from adding non-existent Bestor crystals to water, and therefore, option 4 is selected.

The measurement instrument was administered three times at two-year intervals to the Ss in their own classrooms by their own teachers using standardized procedures. Ss were led to believe that the instrument was an aptitude test and a part of the school's regular testing program.

Results and Discussion

Table 1 presents the Kuder-Richardson Formula 20 (KR20) reliabilities for the cross-sectional data by grade. The internal consistency of the TW results was not high for any grade. However, for a scale of only 16 items the KR20's were comparatively high and appeared to increase over grades with a median KR20 of .40 for grades 5 through 8 and a median KR20 of .65 for grades 9 through 12. The two-and four-year test-retest correlations (see Table 2) are of approximately the same magnitude as the KR20 reliabilities giving indication that TW was a somewhat stable trait over the intervals observed.

Means, standard deviations, and sample sizes by sex are reported in Table 3. Sex by year MANOVA trend analysis resulted in no evidence of stable sex differences or sex by year interactions. Significant ($P<.05$) linear trends were observed over all of the four-year intervals with a significant quadratic trend over the 7 through 11 interval. The proportion of total sums of squares attributable to the linear trend was .41 for the 5, 7, 9 group,

.35 for the 6, 8, 10 group, .44 for the 7, 9, 11 group (.09 attributable to the quadratic component for the 7, 9, 11 group), and .37 for the 8, 10, 12 group. In summary, TW appears to be somewhat stable and increasing over grades with no evidence of sex differences or sex by grade interaction.

Given the present results, the generality of TW across content areas (Millman, 1966), and that TW cues are present in objective items on standardized as well as teacher-made tests, then the following should be considered by persons involved with educational measurements: (a) the degree to which a test measures TW is related to the observed stability of the results of aptitude or achievement tests and (b) if TW is not controlled then examinees low in TW will be penalized on objective test scores across subject areas both at a point in time and over time. Since it is unlikely that TW cues can be totally eliminated from objective tests it is recommended that an attempt be made to control for individual differences in TW by teaching test-taking strategies to all students and especially those identified as low in TW.

References

Crehan, K.D., Koehler, R.A., and Slakter, M.J. Longitudinal studies of test-wiseness. Journal of Educational Measurement, 1974, 11, 209-212.

Diamond, J.J. and Evans, W.J. An investigation of the cognitive correlates of test-wiseness. Journal of Educational Measurement, 1972, 9, 145-150.

Ebel, R.L. Measuring Educational Achievement. (2nd. ed.), New Jersey: Prentice-Hall, 1972.

Gibb, B.G. Test-wiseness as secondary cue responses (Doctoral dissertation, Stanford University) Ann Arbor, Michigan: University Microfilms, 1964. No. 64-7643.

Millman, J. Test-wiseness in taking objective achievement and aptitude examinations. 1966, Final Report, College Entrance Examination Board.

Millman, J., Bishop, H. and Ebel, R. An analysis of test-wiseness. Educational and Psychological Measurement, 1965, 25, 707-726.

Moore, J.C., Schutz, R.E., and Baker, R.L. The application of a self-instructional technique to develop a test-taking strategy. American Educational Research Journal, 1966, 3, 13-17.

Slakter, M.J., Koehler, R.A., and Hampton, S.H. Grade level, sex, and selected aspects of test-wiseness. Journal of Educational Measurement, 1970a, 7, 119-122.

Slakter, M.J., Koehler, R.A., and Hampton, S.H. Learning test-wiseness by programmed texts. Journal of Educational Measurement, 1970b, 7, 247-254.

TABLE 1

KR-20 Reliabilities for Cross-Sectional Observations of TW

	Grade							
	5	6	7	8	9	10	11	12
First Observation	.40	.18	.44	.40	.63	.60	.72	---
Second Observation	.41	.30	.51	.19	.65	.62	.71	---
Third Observation	---	---	---	---	.65	.52	.66	.69

TABLE 2

Test-Retest Correlations (Stabilities)
of Matched TW Observations Over Two- and Four-Year Intervals

		Interval				Interval				Interval				Interval		
		2 Year		4 Year		2 Year		4 Year		2 Year		4 Year		2 Year		4 Year
Sex	N	5-7	7-9	5-9	N	6-8	8-10	6-10	N	7-9	9-11	7-11	N	8-10	10-12	8-12
Male	46	.41	.39	.37	42	.04	.35	.11	37	.42	.60	.50	23	.66	.67	.57
Female	29	.31	.39	.50	34	.14	.35	.37	36	.58	.77	.58	41	.43	.52	.45

TABLE 3

Mean TW by Sex Within Grade for Matched Longitudinal Data (Standard Deviations in Parentheses)

Sex	N	Gd.	First Observation	Gd.	Second Observation	Gd.	Third Observation
Male	46	5	5.93 (2.27)	7	6.74 (2.62)	9	8.22 (3.12)
Female	29		5.69 (2.36)		7.31 (2.45)		8.31 (2.77)
Male	42	6	7.50 (1.76)	8	8.50 (1.89)	10	9.31 (2.79)
Female	34		6.97 (1.75)		7.74 (1.94)		9.29 (2.29)
Male	37	7	6.57 (2.54)	9	8.16 (2.80)	11	8.73 (3.02)
Female	36		6.03 (2.71)		8.22 (3.22)		8.61 (2.90)
Male	23	8	7.61 (2.33)	10	8.61 (2.98)	12	10.26 (3.63)
Female	41		7.56 (2.51)		8.63 (2.95)		9.41 (2.78)